On the Front Lines: Unequal Exposure to COVID-19 Risk and Poor Air Quality in the United States

Author: Daniel Vila

Supervisor: Dr Tom McDermott

National University of Ireland, Galway

J.E Cairnes School of Business and Economics

Head of School: Dr. Tom Acton

A thesis submitted in partial fulfilment for the requirement of degree of MSc Global Environmental Economics

July 2020

Abstract

This paper uses data from over 3,000 counties in the US from the JHU CSSE Coronavirus Resource Center to build a nationwide cross sectional zero-inflated negative binomial model using county level data to model COVID-19 death counts. By controlling for a wide range of confounding factors and creating supporting models for changes within several of these interrelated controls, we create models akin to estimating the difference-in-difference of the effects that COVID-19 has had on workplace mobility, unemployment, and pollutant concentrations. This allows this paper to uniquely discuss the effect of the virus on the US population.

The results demonstrate that there is a positive and statistically significant relationship between NO_2 concentrations and COVID-19 outcomes: A one standard deviation increase in pre-COVID-19 NO_2 concentration is expected to result in 6.5 more COVID-19 deaths. The results also demonstrated that counties with higher percentages of minority populations in the US are disproportionately affected by COVID-19: A county with 10% more Black people is expected to have 4 more COVID-19 deaths. A county with 10% more Hispanic people is expected to have 3 more COVID-19 deaths. And, although responses to COVID-19 (stay at home orders, non-essential business closures, etc.) resulted in decreases in overall NO_2 concentrations, the improvements in air quality were often not present in counties with high percentages of Black, Hispanic, or poor individuals. This indicates that not only have minority populations been disproportionately affected by the immediate impacts of COVID-19 (cases and deaths), but they have also been disproportionately subjected to harmful secondary affects (increases in NO_2 concentrations).

Contents

1 Introduction **1**

2 Literature Review **6**

3 Data and Methodology **10**

4 Results and discussion **21**

5 Conclusion **36**

References **37**

A Variables **41**

B Descriptive Statistics **44**

1 Introduction

COVID-19 is a global pandemic that has reached over 188 countries and territories worldwide. Globally, at the time of writing there have been over 16 million cases and 650,000 deaths (according to the European Centre for Disease Prevention and Control). In the United States (US) alone, there have been over 4.2 million cases and 147,000 deaths (according to the US Center for Disease Control (CDC) and Prevention). The goal of this paper is to examine the disproportionate effects COVID-19 has on minority populations in the US, and to empirically examine how and why this has been the case. This will be accomplished by combining several interrelated models of varying data related to the COVID-19 outbreak in the US. Aside from the direct impact on human health (cases and deaths), there are also grave social and economic impacts that are associated with the virus. According to the United Nations Development Program, COVID-19 will most likely increase poverty and inequality on a global scale. This is particularly concerning for the US because there is more poverty and economic inequality in the US than anywhere else in the industrialized world (Barkan et al. (2011)). US economic inequality has also continued to increase in recent decades. From 1979 to 2005, after adjusting for inflation, the post-tax income of the nation's wealthiest families grew by a much greater amount than that for the poorest families; it grew by only 6% for the poorest fifth of the population, but by 80% for the wealthiest fifth. The top 1% of the nation's families saw their incomes increase by a whopping 228%. The average poverty rate (defined here as households that receive less than half of the national median income after taxes) of Western democracies is 9.8%. The US is 1.73 times greater than this average: 17% of US households lived in

poverty in 2009 (Barkan et al. (2011); Mishel et al. (2009)).

In and of themselves, these numbers are concerning. However, this also presents added concern during the COVID-19 pandemic because the virus has disproportionately affected individuals in lower income distributions in countries around the world. This could potentially be explained by the fact that on average poor people tend to live in tighter housing conditions and with less access to affordable healthcare and medical facilities. However, many of the effects that COVID-19 has imposed on these populations, including the threat of increased famine and reversal of long-term gains against global poverty, are a result of the policy responses to COVID-19 (Abonyi (2020)).

In the US, Black and Hispanic people have had a higher COVID-19 death rate. This has been explained, in part, by the higher prevalence of underlying health conditions (such as diabetes, obesity, and asthma) (Millett et al. (2020)). However, the new CDC data also reveals a significant imbalance in the number of virus cases, not just deaths — a fact that underscores inequities unrelated to other health issues (New York Times, 2020). Other factors contributing to this heightened susceptibility of Black and Latino populations may include insufficient access to optimal medical care, poor nutrition, socioeconomic stresses, and co-exposure to other pollutants such as bioaerosols (American Lung Association (2001)). The underlying conditions that partially explain the disproportionate outcomes are themselves often caused by environmental injustices and racism. In the US, institutional rules, regulations, policies, and/or corporate decisions often deliberately target certain communities for locally undesirable land uses and lax enforcement of zoning and environmental laws (Green Action, 2020). This results in communities being disproportionately exposed to toxic and hazardous waste based upon race and income. This is hardly a new concept; according to a report released after the Urban Air Pollution and Health Inequities Workshop in 1999 (American Lung Association (2001)) (organized by the American Lung Association with major support from the National Institute for Environmental Health Sciences, California Air Resources Board, and CDC) there is a disproportionately higher prevalence of harmful environmental exposures, particularly air

pollution, for certain populations. These populations most often reside in urban settings, have low socioeconomic status, and include a large proportion of ethnic minorities (American Lung Association (2001)). These harmful environmental exposures have led to higher prevalence of underlying health problems in urban and minority populations. For example, the asthma mortality rate among Blacks in the US was 2.5 times higher than for whites; in Chicago, it was 4.7 times higher. Further small-area ecologic analysis showed positive correlations between asthma hospitalizations and higher poverty rate and proportion of nonwhite residents (American Lung Association (2001)).

As is the case for many systemic issues in the US, poverty and inequality are also unequally distributed by race—particularly between white and Black households. According to a report by Hanks et al. (2018), Black households have far less access to tax-advantaged forms of savings, due in part to a long history of employment discrimination and other discriminatory practices. A well-documented history of mortgage market discrimination means that Black people are significantly less likely to be homeowners than whites, which means they have less access to the savings and tax benefits that come with owning a home. Persistent labor market discrimination and segregation also force Black individuals into fewer and less advantageous employment opportunities than their white counterparts. Thus, Black Americans have less access to stable jobs, good wages, and retirement benefits at work. Persistent housing and labor market discrimination and segregation has worsened the damaging cycle of wealth inequality and has left Black households and citizens less equipped to cope with risk and uncertainty.

Evidence of these lingering racial disparities can be clearly seen in COVID-19 outcomes. The virus has disproportionately affected people of color. Early data on viral outcomes indicated that Black and Hispanic people were being harmed by the virus at higher rates than their white counterparts (Millett et al. (2020); New York Times, 2020). New federal data from the CDC is starting to reveal an even clearer picture: According to recent reporting from the New York Times, Black and Hispanic people have been disproportionately affected by COVID-19 in a widespread manner that spans the country,

throughout hundreds of counties in urban, suburban and rural areas, and across all age groups (New York Times, 2020). Hispanic and African-American residents of the United States have been three times as likely to become infected as their white neighbors. And Black and Hispanic people have been nearly twice as likely to die from the virus as white people (New York Times, 2020). The new CDC data, which gives detailed characteristics of 640,000 infections detected in nearly 1,000 U.S. counties, also showed several pockets of disparity involving Native American people. In much of Arizona and in several counties in other states, they were far more likely to become infected than white people. For people who are Asian, the disparities were generally not as large, though they were 1.3 times as likely as their white neighbors to become infected. This data was only made publicly available after the New York Times sued the CDC. It is also interesting to note that since it has been released to the public, the Trump administration has ordered hospitals to bypass the CDC and send all COVID-19 patient information to a central database in Washington (New York Times, 2020). As we continue to investigate how and why racial disparities have persisted in American society, it is important to acknowledge the political context that has lead us to this point. A more specific and nuanced discussion of American political history is needed to fully understand the role that the state has played in shaping life for Black Americans. However, this paper will focus on the specific mechanics that have caused the aforementioned disproportionate effects of the COVID-19 pandemic.

Other COVID-19 confounders include that minority populations live in tighter quarters and are less likely to be able to work from home, or take off work altogether. Across the country, 43 percent of Black and Hispanic workers are employed in service or production jobs that for the most part cannot be done remotely. Only about one in four white workers held such jobs. Also, Hispanic people are twice as likely as white people to reside in a crowded dwelling — less than 500 square feet per person (according to the American Housing Survey) (New York Times, 2020).

The national figures for infections and deaths from the virus understate the black-

white disparity to a certain extent. The virus is far more prevalent among older Americans, who are disproportionately white compared with younger Americans. When comparing infections and deaths within specific age cohorts, the disparities are even more extreme. The new CDC data shows that Hispanic people between the ages of 40 and 59 have been infected at five times the rate of white people in the same age cohort. The differences are even more stark when it comes to deaths: Of Hispanic people who died, more than a quarter were younger than 60. Among white people who died, only 6 percent were that young (New York Times, 2020).

The goal of this paper is not only to examine the disproportionate effects of COVID-19 on minority populations in the US, but also to empirically examine the specific mechanics that have given rise to these outcomes. This will be accomplished by focusing on county level data for concentrations of air pollution, unemployment, workplace mobility, and a number of other socio-economic characteristics. By controlling for a wide range of confounding factors in each of these inter-related models, we are able to uniquely discuss the effect of the virus on the US population by a number of different factors including race, income, and air quality.

2 Literature Review

Recent studies have demonstrated a clear and statistically significant positive relationship between local air pollution and COVID-19 outcomes (Wu et al. (2020); Cole, Ozgen & Strobl (2020); Travaglio et al. (2020); Conticini et al. (2020)). Other studies have identified the significant improvements in air quality that have resulted from Covid-19 lockdowns (Cicala et al. (2020); Cole, Elliott & Liu (2020)). It may seem counterintuitive to claim that air pollution worsens the outcome of a virus that has led to improvements in air quality. However, according to the relevant and forthcoming literature, that appears to be what has happened.

It is well known that long term exposure to pollutants such as nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and fine particulate matter (PM2.5) contributes to cardiovascular disease, reduces lung function, and causes respiratory illness (Cole, Ozgen & Strobl (2020)). These pollutants have been shown to cause a persistent inflammatory response even in the relatively young, and to increase the risk of infection by viruses, like COIVD-19, that target the respiratory tract (Cole, Ozgen & Strobl (2020), Travaglio et al. (2020); Conticini et al. (2020)). So, even if the economic shut-down associated with the virus has led to significant improvements in air quality in certain places, areas that are associated with higher levels of historical pollutant concentrations are still associated with worse COVID-19 outcomes.

Travaglio et al. (2020) finds evidence of a correlation between Covid-19 cases and concentrations of nitrogen oxides and ozone, while Ogen (2020) examines 66 regions across Italy, France, Germany, and Spain and finds similar evidence. Conticini et al. (2020) et al.

focus on Northern Italy and conclude that pollution concentrations are a likely contributor to the high Covid-19 death rates experienced in that region. A preliminary analysis of the Netherlands also finds evidence of a link between concentrations of PM2.5 and Covid-19 cases (Andree (2020)).

Cole, Ozgen & Strobl (2020) examined long term air pollution exposure and Covid-19 in 355 municipalities in the Netherlands. Using detailed secondary and administrative data they find compelling evidence of a positive relationship between air pollution, and particularly PM2.5 concentrations, and Covid-19 cases, hospital admissions and deaths. This relationship persists after controlling for a wide range of explanatory variables. Their results indicate that a 1 $\mu/[m^3]$ increase in PM2.5 concentrations is associated with 9.4 more Covid-19 cases, 3.0 more hospital admissions, and 2.3 more deaths. This analysis is unique as parts of the Netherlands are associated with air pollution due to intensive livestock farming as opposed to urban environments. In terms of COVID-19 it is important to distinguish between the effects of air pollution and population density. While population density is certainly a contributing factor in person to person transmission, research suggests that the virus is more deadly in areas with poor air quality. The Cole, Ozgen & Strobl (2020) findings on air quality may prove particularly beneficial for rural areas where healthcare infrastructure and coordination may be less developed.

Using data from 3,000 counties in the United States (representing 98% of the population) from Johns Hopkins University, Center for Systems Science and Engineering (JHU CSSE) Coronavirus Resource Center, Wu et al. (2020), put together a nationwide cross sectional negative binomial model using county level data to model COVID-19 death counts controlling for 20 potential confounding factors including population size, age distribution, population density, time since the beginning of the outbreak, time since state's issuance of stay-at-home order, hospital beds, number of individuals tested, weather, and socioeconomic and behavioral variables such as obesity and smoking. This is significant to this report because it uses the same county level US data (JHU CSSE). A few key differences will be presented in the following section that specifically highlights the data

and methodology of this report. However, Wu et al. (2020) found that an increase of only 1 μ/m^3 in PM2.5 is associated with an 8% increase in the COVID-19 death rate (95% confidence interval [CI]: 2%, 15%). The results were statistically significant and robust to secondary and sensitivity analyses. They concluded that a small increase in long term exposure to PM2.5 is associated with a large increase in the COVID-19 death rate.

Coven & Gupta (2020) investigate the role of demographic differences in mobility in explaining disparities in COVID-19 outcomes in New York City. The authors reflect that the disproportionate disease burden of Black and Latin New Yorkers may reflect long-standing disparities in health outcomes in these communities, but it may also reflect novel patterns of mobility and exposure with respect to the new COVID-19 infection. They conclude that richer and younger neighborhoods see far greater increases in the propensity of individuals to leave the city, starting around March 14, 2020. They also find differential patterns of sheltering in place during the pandemic in low-income, Black, and Hispanic neighborhoods. This is the case during normal work hours, when it is plausible that members of these populations are likely to be frontline workers, while other populations are more easily able to work remotely. The same is true during non-work hours, when adherence to shelter in place orders may depend on the frequency of visits to retail establishments instead of ordering food and groceries delivery services, or when essential workers might be required to work longer hours. These findings complement other well-established health disparities which may impact the severity of COVID-19 for different populations.

It is important for the goals of this paper that the information presented above on the relationship between socio-economic factors, COVID-19, and pollution is all presented within an economic context. Graff Zivin & Neidell (2012) offer an overview on the contributions that economists have made towards understanding the relationship between the environment and individual human health. They highlight the role that optimizing behavior (such as residential sorting by income) can lead to a non-random allocation of pollution within an economy. They also discuss how environmental conditions has allowed

for this type of analysis to expand its focus to include measures of human capital such as labor supply, productivity, and cognition. They offer a litany of quasiexperimental economic research to support their argument, and conclude that environmental regulation may actually contribute to economic productivity and thus fruitfully be viewed as an investment in economic growth. And, since exposure to poor environmental quality within and across countries tends to correlate with low income, their results point toward a new form of poverty trap. With this in mind, this paper will further explore the roles that race (and other socio-economic factors) and COVID-19 play in contributing towards this "new" poverty trap.

In this paper, I explore the relationship between COVID-19 outcomes and air pollution that has been discussed in the nascent literature (mentioned above). The unique contribution to the discourse on this topic presented here is not only to address the connection between COVID-19 and air pollution in the US, but to also to examine how this has resulted in minority populations being disproportionately impacted by COVID-19 outcomes. This will be accomplished by using data from over 3,000 counties in the US from the JHU CSSE Coronavirus Resource Center to build a nationwide cross sectional study using county level data to model COVID-19 death counts. This is the same data used by Wu et al. (2020). However, the model I built presents two key improvements: First, as this analysis happened after theirs, I analyzed more data from a longer period of the COVID-19 pandemic. Second, the methodology I used was slightly different from theirs, and as I will argue in the next section, represents a marginal improvement in how it fits the data. Also, by combining this model of COVID-19 deaths with other interrelated models of google workplace mobility (similar to Coven & Gupta (2020) but expanded to the entire US), changes in unemployment, and changes in pollutant concentrations, I am uniquely able to not only analyze the outcomes of COVID-19, but also begin to empirically analyze how and why these outcomes have disproportionately affected minority populations in the US.

3 Data and Methodology

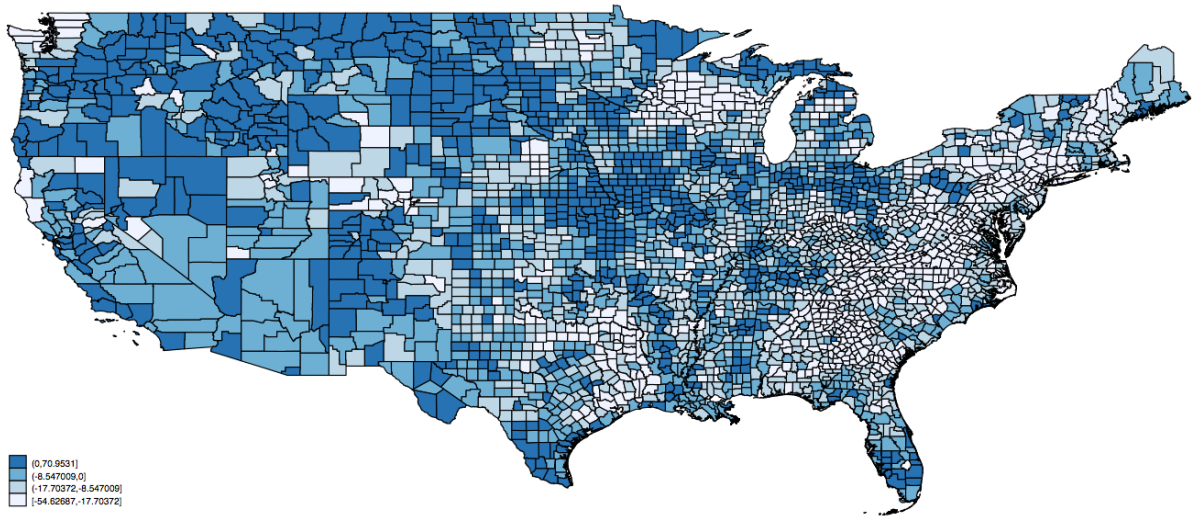
The data for this project was compiled from a number of different sources. The COVID-19 infection intensity data (deaths, confirmed cases, hospitalizations, beds_icu) comes from the Johns Hopkins University Center for Systems Science and Engineering dataset (JHU CSSE) (Dong et al. (2020)). However, it is important to note that this dataset also contains data from a number of sources including the CDC (and the European CDC), World Health Organization, and various state level departments of health.

The pre-pandemic, pandemic, unemployment, and policy control variables (percent_black, percent_hisp, povprop, percent_over_65, popdensity_2018, rmortality_2018, uninsured_2017, and unem_trend) come from the Johns Hopkins Population Center (HPC). Drawing from trusted sources (including the US census, American Community Survey, CDC, and the Bureau of Labor Statistics) the HPC Data Hub offers US county-level data on COVID-19 and sociodemographic data necessary for population-based social science, epidemiological, medical and public health research to provide evidence-based policy recommendations for curbing the pandemic. Data on the timing of state level COVID-19 policy comes from the Boston University COVID-19 US State policy database (CUSP) (Raifman et al. (2020)).

The mobility data used in this project comes from Google's COVID-19 Community Mobility Reports. These reports aim to provide insights into what has changed in response to policies aimed at combating COVID-19. The reports chart movement trends over time by geography, across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential. They are presented

as daily percent changes in mobility from the pre-COVID baseline. For the purpose of this project, I calculated the mean change in mobility for each county during the COVID-19 pandemic.

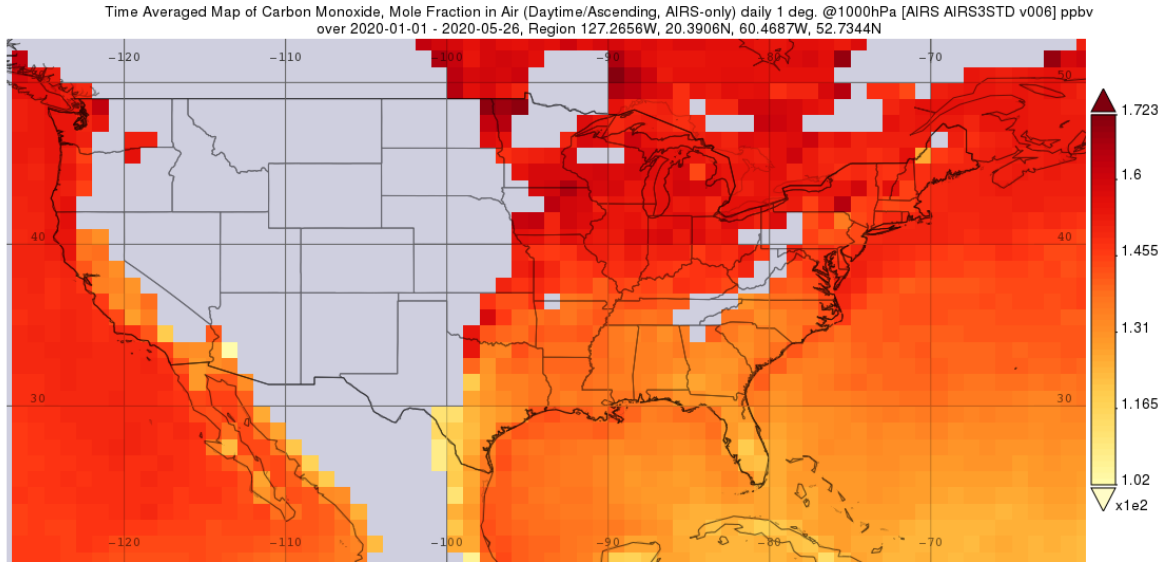
Analyses and visualizations of pollutant concentrations used in this project were produced with the Giovanni online data system, developed and maintained by the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC). Initially, I wanted to control for four pollutants: Nitrogen Dioxide, Sulphur Dioxide, Particulate Matter (PM2.5), and Carbon Monoxide. Using the Giovanni tool, I downloaded two (2019 and 2020) time averaged maps for each of these pollutant concentrations from 1 January – 10 June (as an estimate of the concentrations during the pandemic). I then uploaded these maps into QGIS and overlaid them with a map of US county boundaries. I was then able to calculate the 2019 mean concentration, 2020 mean concentration, and percent change for each pollutant by county. See the map below for the average change in NO_2 concentrations for each county.



For the concentration of nitrogen dioxide, I used the Giovanni parameter NO_2 Tropospheric column (30% cloud screened) from the Ozone Monitoring Instrument (OMI) aboard the Aura satellite. This instrument measures the number of molecules of NO_2 in an atmospheric column (from the Earth's surface to the top of the atmosphere) above a

square centimeter of the surface with a spatial resolution of 0.25°. NO_2 data is provided only for near clear sky conditions (i.e only those NO_2 retrievals are used in the analysis for which Cloud radiance fraction is less than 30%).

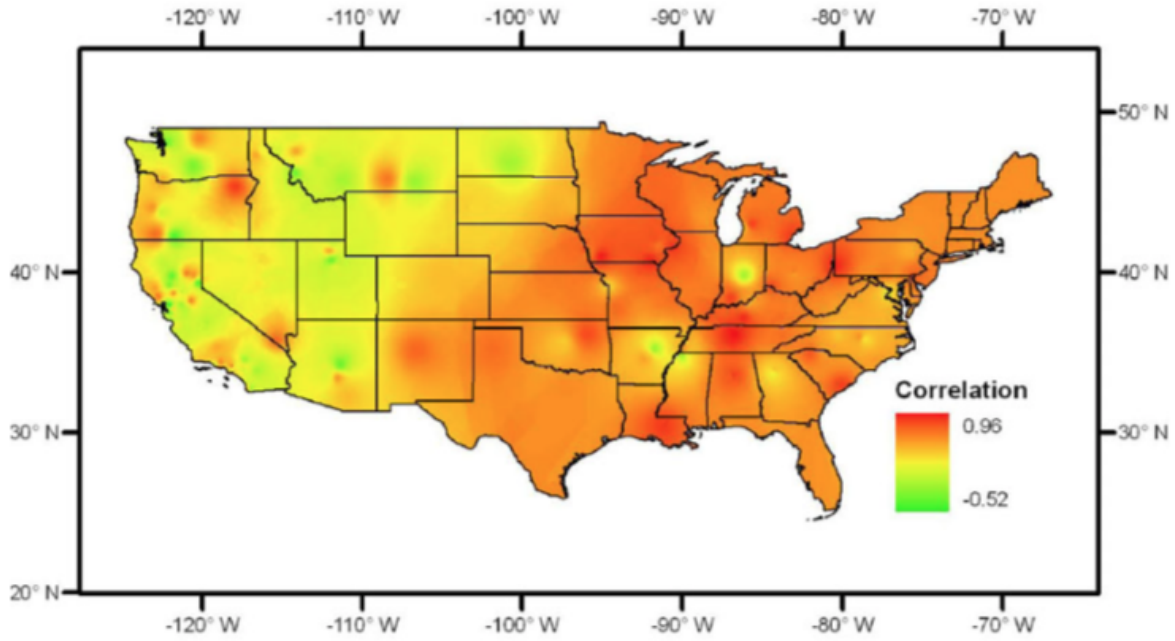
For the concentration of carbon monoxide, I used the parameter of Mole Fraction in Air (Daytime/Ascending) from the Atmospheric Infrared Sounder (AIRS) aboard the Aqua satellite. This gave a daily value of the amount (in ppbv) of CO in the atmosphere during daytime hours at 1°of spatial resolution. The concentration of carbon monoxide (CO). As a total column measurement, it is the number of molecules of CO in an atmospheric column from the Earth's surface to the top of the stratosphere above a square centimeter of the surface. As a profile, or layer measurement it is the volume mixing ratio in parts per billion (ppbv). CO is a poisonous, odorless, and colorless gas. It is produced by incomplete combustion of fossil fuels, biomass burning, and the oxidation of methane and other hydrocarbons. CO is not a direct greenhouse gas; however, CO and other pollutants can affect tropospheric ozone, carbon dioxide, and methane. Thus, CO can have an indirect effect on climate. CO is a relatively short-lived species, with a lifetime on the order of two months in the troposphere. Because of its relatively short lifetime, CO is not well-mixed, and thus it is often used as a tracer to track polluted air masses in the troposphere. Unfortunately, the measurements of CO over the continental US were characterized by large amounts of missing data (see map at top of next page).


Time Averaged Map of Carbon Monoxide, Mole Fraction in Air (Daytime/Ascending, AIRS-only) daily 1 deg. @1000hPa [AIRS AIRS3STD v006] ppbv
over 2020-01-01 - 2020-05-26, Region 127.2656W, 20.3906N, 60.4687W, 52.7344N
1.723
1.6
1.455
1.31
1.165
1.02
x1e2

For the concentration of sulphur dioxide, I used the parameter of SO_2 Column Mass Density (ENSEMBLE) from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) model. The GEOS General Circulation Model (GCM) used in MERRA-2 has been used to generate a 10-member ensemble of simulations, configured following the convention of the Atmospheric Model Intercomparison Project (AMIP; Gates (1992)). This shows the atmospheric concentration of sulfur dioxide. It occurs naturally as a component of volcanic emissions, and is also produced by human activity, mainly the burning of fossil fuels containing sulfur. Unfortunately, this measurement was highly correlated with NO_2 concentrations, and was subsequently dropped from the analysis.

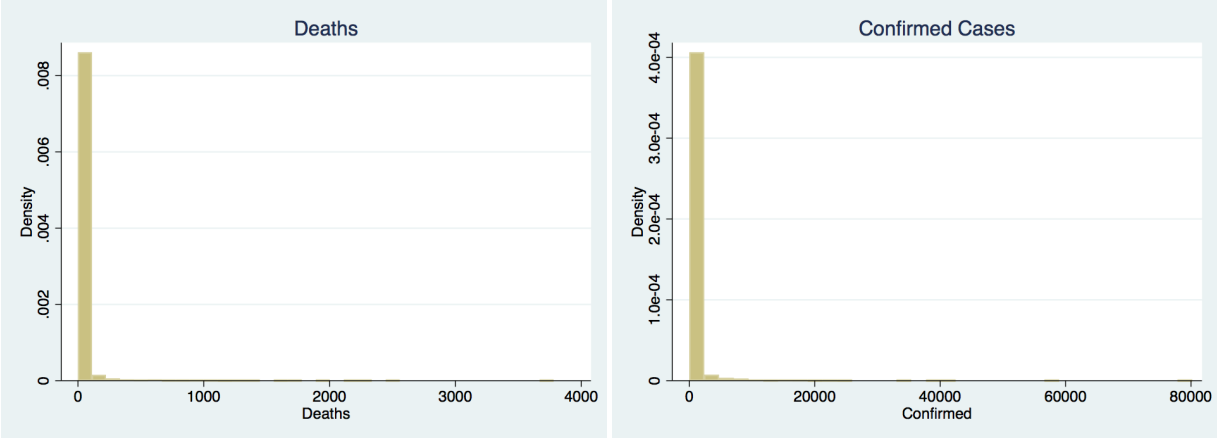
In order to estimate PM2.5 concentrations, I used the parameter of Aerosol Optical Depth (AOD) (342.5 nm) Ozone Monitoring Instrument (OMI) Aerosol Product (OMAERO) aboard the Aura satellite, which has a spatial resolution of 0.25°. Aerosol optical depth is the degree to which aerosols prevent the transmission of light by absorption or scattering of light. The aerosol optical depth (AOD) or optical thickness (τ) is defined as the integrated extinction coefficient over a vertical column of unit cross section. It is important to note that studies have found that correlation between PM2.5 and AOD is not spatially consistent due to variation in terrain, land cover, selection of

aerosol model in the AOD retrieval algorithm and meteorological factors (Engel-Cox et al. (2004); Hu (2009)). Hu (2009) calculates a Pearson's correlation value between US EPA Air Quality System monitoring site (ground level measurements of PM2.5) and MODIS AOD satellite measurements. The Pearson's correlation analysis examines the temporal relationship between PM2.5 and AOD for each site. To visualize the spatial variation of the correlations, inverse distance weighted interpolation was used to create a correlation surface. Values for the continental US can be seen in the image below (top of next page). In order to remedy this, studies often calculate exposure to PM2.5 via an established exposure prediction model. The PM2.5 exposure levels are estimated by combining satellite, modeled, and monitored PM2.5 data in a geographically weighted regression. The AOD measurements used in this analysis were not amended in this manner. As such, they can be interpreted as a moderately good proxy for PM2.5 (due to the higher levels of correlation in areas with high populations seen in map below). Due to this issue with AOD measurement, the missing concentrations of CO, and the highly correlated values of SO_2, NO_2 was chosen as the most suitable metric of air quality.



The best measure of COVID-19 intensity in the data was either deaths or confirmed cases, both of which constitute count variables. When building models of count data, it is imperative to test for overdispersion and excess zeros in order to choose the model that best fits the data. Poisson models assume the equidispersion property. That is that the mean of y given x, is equal to the variance of y given x. Looking at the summary statistics of the potential dependent variables available from my data, we can clearly see from the data that the equidispersion property is clearly violated. The mean of confirmed COVID-19 cases per county is 555, and the variance is 7,840,031. The mean amount of COVID-19 deaths per county is 29, and the variance is 25,508. The data clearly exhibits overdispersion, and a Poisson model would not fit.

The simplest way to determine if the data is characterized by excess zeros is to look at the histogram of the dependent variable. Looking at the histogram of confirmed cases and deaths (below), we can see that excess zeros may indeed be a problem. Taking a closer look, we can see that there are 31 observations with zero cases (just over 1% of the 2,976 total observations) and 1,223 observations with zero deaths (41%).



Both variables are highly overdispersed. However, only deaths appear to exhibit excess zeros. Because of overdispersion, I decided against using a Poisson model for either variable. The choice was then between a negative binomial model or a zero-inflated negative binomial model and which variable I wanted to include as the dependent variable. Due to the fact that the data did not exhibit excess zeros I reasoned that a negative binomial model would be appropriate for modeling the confirmed cases. However, because COVID-19 is characterized by large amounts of mild non-life threatening infections (particularly in younger populations), I decided that building a model of COVID-19 deaths would be a more accurate representation of the intensity of the virus. As such, I wanted to seriously examine building a zero-inflated negative binomial model (ZINB) of COVID-19 deaths.

ZINB models are a combination of two different models: a binary model of the zero distribution, and a count model of the rest of the non-zero data. One question that is central in specifying models for this type of data is whether the zero and positive observations are generated by the same mechanism or whether the zeros are somehow different (Silva et al. (2015)). If the zeros are believed to be generated by a different process, the covariates are allowed to affect the conditional distribution of y in two different ways, leading to a double-indexed model (Silva et al. (2015)). The two different processes that would generate an observation of 0 in my data are (1) the population (in this case county)

has not been exposed to COVID-19, or (2) the population has been exposed to COVID-19, but the intensity of the exposure is small enough that it has not resulted in a single death, as was the case for 1,192 counties. Both the ZINB and negative binomial models define the conditional expectation of y given a set of covariates x. Therefore, according to Silva et al. (2015), the suitability of each of these models can be gauged by testing the corresponding conditional expectation against that of the alternative. Doing so would allow us to determine if the estimate of the conditional expectation of y obtained using the ZINB model can be used to improve the prediction of y obtained under the normal negative binomial model. If this is the case, there would then be evidence against using the negative binomial model because this would imply that the model is not explaining some features of the data that are captured by the ZINB (Silva et al. (2015)).

Historically, economists have used the Vuong (1989) test in order to test for zero-inflation. However, recent literature (Silva et al. (2015); Wilson (2015)) suggests that this test is inappropriate because it was designed for non-nested models. Wilson (2015) suggests that this is due to a misunderstanding of what is meant by the term "non-nested model". Vuong (1989) specifically relies on the distribution of the log-likelihood ratios of non-nested models being normal, which is dependent on six assumptions that refer to various topological and measure theoretic properties which are unlikely to be understood without a formal background in mathematics. According to Wilson (2015), zero-inflated models are characterized by a non-normality of the sampling distribution of the zero-inflation parameter. As Vuong's subsequent theoretical development of the distributions of log-likelihood ratios depends on this normality, his theory is not applicable to test for zero-inflation, and doing so results in a bias towards adopting the zero-inflated model.

Silva et al. (2015) provide a simple regression based specification test (HPC test)

that can be used to test for zero-inflation. In the context of this data, the test would be:

$$\begin{aligned} H_0 : E[y|x] &= \exp(x'\beta) \\ H_A : Pr(y > 0|x) &= \phi(x'y) \\ E(y|x, y > 0) &= \exp(x'\beta) \end{aligned}$$

The HPC test checks whether the errors of the model under the null have zero expectation when the weight given to each observation depends on the difference between the conditional expectations of the two models (Silva et al. (2015)). Silva et al. (2015) claims that this test outperforms others because it partially accounts for the presence of heteroskedasticity and consequently it is better suited to test for zero-inflation.

In order to test which model performed best I put together both a zero-inflated and a negative binomial model for COVID-19 deaths. The ZINB model incorporated an inflation model using a probit regression controlling for confirmed cases to predict the probability of 0 deaths. After this inflation model is complete a negative binomial model is then estimated with the remaining data. The HPC test was used to compare this ZINB model with the negative binomial alternative. Both regressions controlled for the same independent variables: the standardized pre-COVID-19 (2019) concentrations of AOD, NO2, and CO, due to the nascent literature that suggests a positive relationship between air pollution and COVID-19; percentage of population that is Black and Hispanic, due to the higher prevalence of air pollutants in minority neighborhoods in the US; percentage of population living below the federal poverty line pre-COVID-19 (2018) due to the fact that globally pollution levels and income are inversely related geographically; number of intensive care unit hospital beds due to the fact that one of the main concerns about the COVID-19 outbreak is overwhelming local hospital capacity; percentage of the population over 65 years of age due to the fact that COVID-19 has a much higher death rate for older populations; crude death rate of respiratory diseases pre-COVID-19 (2018)

to further explore the relationship between COVID-19 and areas with higher historical pollution concentrations; and the proportion of the population without health insurance pre-COVID-19 (2018) to help shed light on financial risks that COVID-19 could impose on individuals.

The HPC test was conducted with the negative binomial model as the null hypothesis and the ZINB model as the alternative hypothesis. The output of the test was: t = 13.96 and prob > t = 0.000. Based on this I rejected the null hypothesis in favor of the ZINB model. The negative binomial model produced an R^2 value of 0.1503 (calculated as the square of the correlation between the dependent variable and the estimated coefficient), while the ZINB model produced an R^2 value of 0.1608. This indicates that the ZINB model fits the data marginally better than the negative binomial model. I argue that this improvement most likely comes from including the confirmed cases of COVID-19 in the inflation model. While a negative binomial model of deaths does not include any of the data on cases unless it is included as a control (which would bias the results due to the fact that deaths and cases were highly correlated: 0.94). The results of this model can be found in the following section.

The other regressions used in this analysis are ordinary least squared regressions designed to help explain some of the results of the ZINB model of COVID-19 intensity. The dependent variables for these models include: changes in workplace mobility; changes in unemployment; the ratio of deaths to cases (deaths/cases); the change in NO2 concentrations during the shut-down. Three of these models employ dependent variables that are changes that have been a result of the COVID-19 outbreak in the US. By modeling these changes across counties with varying COVID-19 intensity, these models are akin to a difference-in-difference estimation of the causal effects that COVID-19 has had on workplace mobility, unemployment, and pollutant concentrations. This does not mean that this paper is estimating a causal relationship. However, without individual data on COVID-19 outcomes and the relevant confounding factors, this may be one of the best ways to study the effect COVID-19 is having. These models are intended to estimate the

effects on the entire population. As such, counties with larger populations should have more weighting in the calculation of these effects. In order to make these models more representative of the effects on the population in general, each of these regressions was analytically weighted according to the total population.

4 Results and discussion

The tables below contain the results from the ZINB model. Note that this model was estimated three times: once including variables for pre-COVID-19 pollution concentrations, once without controlling for pollution concentrations, and finally once in an incredibly limited model isolating the effects of poverty and uninsured individuals on deaths. The variable "confirmed" in each of the inflation models represents the binary—in this case probit—aspect of the ZINB model is used to determine the probability that a county would certainly have zero deaths. The negative and statistically significant coefficient on this variable indicates that as the number of confirmed COVID-19 cases rises in a county, the less likely it becomes that said county will have zero deaths. While the magnitude of the coefficient (-0.0647 for the pollution model and -0.0660 for the pollution free model) is certainly low, it is important to remember that there were 1,192 counties that had confirmed cases with no deaths.

	(1)	(2)	(3)	(4)
				Marginal
VARIABLES	ZINB	ZINB	ZINB	Effects
AOD_2019_norm	-0.0273			-0.305
	(0.0453)			(0.505)
CO_2019_norm	-0.0963*			-1.076*
	(0.0543)			(0.607)
NO2_2019_norm	0.585***			6.532***
	(0.0367)			(0.452)
percent_black	0.0362***			0.405***
	(0.00357)			(0.0402)
percent_hisp	0.0267***			0.299***
	(0.00396)			(0.0449)
pov_percent	-0.0514***			-0.574***
	(0.00787)			(0.0882)
beds_icu	0.00615***			0.0686***
	(0.000476)			(0.00568)
percent_over_65	0.0333***			0.372***
	(0.00598)			(0.0668)
popdensity_2018	-0.000172***			-0.00192***
	(5.88e-05)			(0.000658)
rmortality_2018	-0.00255***			-0.0285***
	(0.000982)			(0.0110)
uninsured_per_2017	-0.121***			-1.346***
	(0.0134)			(0.157)
confirmed		-0.0647***		
		(0.01000)		
lnalpha			0.397***	
			(0.0415)	
o.confirmed				-
Constant	3.070***	1.189***		
	(0.204)	(0.155)		
Observations	1,853	1,853	1,853	1,853

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
NOTE: All ME at mean value

VARIABLES	(1) ZINB	(2) ZINB	(3) ZINB	(4) Marginal Effects	(5) Marginal Effects
percent_black	0.0327***			0.393***	
	(0.00312)			(0.0374)	
percent_hisp	0.0234***			0.281***	
	(0.00403)			(0.0485)	
pov_percent	-0.0456***			-0.547***	-1.558***
	(0.00601)			(0.0738)	(0.235)
beds_icu	0.00658***			0.0791***	
	(0.000617)			(0.00764)	
percent_over_65	0.0379***			0.455***	
	(0.00651)			(0.0795)	
popdensity_2018	0.00126***			0.0152***	
	(0.000194)			(0.00234)	
rmortality_2018	-0.00317***			-0.0381***	
	(0.000987)			(0.0122)	
uninsured_per_2017	-0.141***			-1.690***	-5.296***
	(0.0114)			(0.147)	(0.443)
confirmed		-0.0660***			
		(0.00666)			
lnalpha			0.674***		
			(0.0365)		
o.confirmed				-	-
Constant	3.342***	1.340***			
	(0.181)	(0.102)			
Observations	2,476	2,476	2,476	2,476	2,975

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
NOTE: All ME at mean value

One of the proposed partial explanations for why minority groups are suffering more from COVID-19 has been that minority populations are often subjected to higher levels of air pollution (The Guardian, 2020); (American Lung Association (2001)). Comparing the

coefficient on the percentage of Black and Hispanic variables in both regressions should help determine how much the effect of race is absorbed by the inclusion of air quality in the regression. Both variables on race exhibit relatively strong, positive, and statistically significant effects on COVID-19 deaths. Both of these effects were also stronger in the regression that included the air quality measures, so the notion that including air quality in this analysis would absorb some of the effect that race has had on COVID-19 does not appear to hold up. According to this analysis (and controlling for pollution concentrations), holding all other marginal effects at their mean value, a county with 10% more Black people is expected to have 4 more COVID-19 deaths. A county with 10% more Hispanic people is expected to have 3 more COVID-19 deaths.

As discussed above, CO and AOD were of little significance practically or economically. If this project was to be repeated, I would like to include more information on air quality to be able to distinguish more specifically what conditions contribute to more intense COVID-19 outcomes. Especially helping to distinguish which pollutants are specifically associated with COVID-19 impacts. This would help inform which communities are more at risk for harmful effects as COVID-19 continues to spread across the US. However, for the sake of this analysis the best metric we have for air quality (in this case air pollution) is the pre-COVID-19 average NO_2 concentration. Holding all the other marginal effects at their mean value, we can see from this output that a county with one standard deviation increase in historical NO_2 concentration is expected to have 6.5 more COVID-19 deaths.

This result is incredibly alarming. Especially considering that according to the American Lung Association et al. (2019) State of the Air report, more than 4 in 10 Americans live with unhealthy air quality that places their health and lives at risk. The report found that 141.1 million people lived in counties with unhealthy levels of either ozone or particle pollution, an increase of more than 7.2 million Americans since the last annual report (2015). In response to the COVID-19 pandemic, on Thursday, 26 March 2020 the US Environmental Protection Agency announced a sweeping relaxation

of environmental rules, allowing power plants, factories and other facilities to determine for themselves if they are able to meet legal requirements on reporting air and water pollution (Wu et al. (2020)).

If this analysis was to be repeated or continued, more attention should be given to determining the specific pollutants that drive COVID-19 intensity. Existing literature (specifically Wu et al. (2020), and Cole, Ozgen & Strobl (2020)) has found a strong and statistically significant relationship between PM2.5 and COVID-19 outcomes (Cole, Ozgen & Strobl (2020) also found less significant positive relationships between COVID-19 outcomes and NO_2 and SO_2). This project found a strong and statistically significant relationship between COVID-19 deaths and NO_2 concentrations. The coefficient on the variable for the pre-COVID-19 (2018) death rate (per 100,000) due to respiratory diseases is not intuitive. Intuitively, if exposure to air pollution is linked to many of the comorbidities that have been associated with poor prognosis and death in COVID-19 patients, including cardiovascular and lung disease, and air pollution exposure is known to cause inflammation, cellular damage, and may suppress early immune response to infections such as COVID-19, then there should be a positive relationship between respiratory disease death rates and COVID-19 deaths (Wu et al. (2020)). This analysis shows the opposite. In both the model controlling for pollution concentrations and the model without pollution controls, the 2018 respiratory disease death rate was negatively (and statistically significantly) associated with COVID-19 deaths. One explanation for this could be that areas with high respiratory disease death rates took COVID-19 more seriously and exhibited stricter social distancing measures. Individual level data on COVID-19 outcomes would help shed more light on this potential explanation. However, due to the nature of COVID-19 individual level data will not be available for the foreseeable future.

Another unexpected finding from these models is the effect that the percent of people below the federal poverty line in 2018 (pov_percent) and the percent of population without health insurance in 2017 (uninsured_per_2017) had on deaths. My hypothesis was that poorer people would be less able to work from home and abide by social distancing.

Also, poorer people tend to live in tighter more crowded living conditions that should put them at higher risk for COVID-19. Both of which should in theory be linked with more intense COVID-19 outcomes. However, the data indicates otherwise. In the model controlling for pollution concentrations, holding every other marginal effect at its mean, a county with 10% more people below the poverty line is expected to have 5 fewer deaths. A county with 10% more uninsured people is expected to have 13 fewer deaths. Both of these effects were statistically significant at the 0.01 level. One explanation for the relationship between COVID-19 deaths and uninsured individuals is the same logic as the explanation for the case of respiratory disease death rate; Uninsured people may have taken the virus more seriously sooner and exhibited stricter social distancing measures not only to avoid the physical effects of COVID-19, but also to avoid the financial burden a strong infection could place on them or their families. While this explanation is plausible, there is no way to test it with this data. My initial reaction to the negative relationship between deaths and poverty was to assume that other variables in the model were absorbing the positive effect that poverty should be having on deaths. Specifically, I thought that the variables controlling for pollution concentrations, race, and population density would be correlated with higher levels of poverty, and when they were removed from the model, poverty would exhibit a positive relationship with COVID-19 deaths. This was not the case. In the final column, we can see that in a model controlling for just poverty and uninsured rates, both exhibit a strong, negative, and statistically significant relationship with COVID-19 deaths. Even though these (and the rest of the control) variables seem to diminish the magnitude of both of these coefficients, there is no evidence from this model that suggests a positive relationship between COVID-19 deaths and higher rates of poverty.

One of the most interesting findings from this analysis is the effect that population density has on COVID-19 deaths. The intuition behind this effect is clear; COVID-19 is transmitted person to person, so areas with higher population densities should be associated with worse outcomes (more deaths) (Rocklöv & Sjödin (2020)). The coefficient on

population density follows this logic in the regression without pollution concentrations. Holding all the other marginal effects at their mean, having 100 more people per square mile is expected to increase COVID-19 deaths by 1. However, when we control for pollution, the sign on this coefficient changes from positive to negative. Having 1000 more people per square mile is expected to lower COVID-19 deaths by 1. This result indicates that pollution concentrations may be better suited to explain COVID-19 intensity than population density. This also supports the Cole, Ozgen & Strobl (2020) finding that rural areas with high levels of air pollution are at high risk for future COVID-19 outbreaks due to the fact that their healthcare infrastructure and coordination may be less developed while high levels of air pollution (from intensive farming for example) may still leave their populations at higher risk for COVID-19. This result is important because population density and air pollution are often correlated (2019 NO_2 concentrations and 2018 population density were 37% correlated in the data used in this regression). Even though it may appear that population density is a driving factor behind COVID-19 intensity, this data suggests that in the US, other factors (including race and air quality) are more significant.

The goal of this project is not only to highlight the disproportionate effects that COVID-19 has had in the US, but also to begin investigating empirically why this has been the case. In order to accomplish this, I put together several ordinary least squares regression models examining several interrelated outcomes that will hopefully help shed some light on this complicated issue. The results of these models can be found in the table below.

VARIABLES	(1) workplace_mean	(2) unem_trend	(3) deaths/confirmed	(4) NO2_trend_std
NO2_2019_norm			0.00501***	
			(0.00108)	
AOD_2019_norm			0.000310	
			(0.00120)	
CO_2019_norm			0.000219	
			(0.00158)	
unem_trend	-0.594***		-0.00127	0.0856***
	(0.0758)		(0.000994)	(0.0186)
percent_black	0.0250***	-0.0113***	-8.19e-05	-0.0103***
	(0.00848)	(0.00252)	(1.00e-04)	(0.00210)
percent_hisp	0.0321***	-0.00271	-0.000311***	0.00616**
	(0.00780)	(0.00208)	(9.63e-05)	(0.00267)
pov_percent	0.186***	0.0356***	0.000417	0.0441***
	(0.0223)	(0.00593)	(0.000266)	(0.00579)
beds_icu	0.00104***		2.76e-06	0.000974***
	(0.000301)		(1.99e-06)	(0.000128)
percent_over_65	0.0225	0.0101	0.000776***	0.0310***
	(0.0248)	(0.00693)	(0.000276)	(0.00455)
popdensity_2018	-0.00205***		-4.99e-06***	-0.000679***
	(0.000132)		(1.36e-06)	(6.80e-05)
rmortality_2018	0.0677***		-1.52e-06	-0.00209***
	(0.00342)		(4.02e-05)	(0.000779)
uninsured_per_2017	-0.0202	0.0248**	-0.000645	0.0139
	(0.0329)	(0.0103)	(0.000404)	(0.0106)
closed_business_duration	-0.0610***	0.0195***	0.000373**	
	(0.0126)	(0.00396)	(0.000161)	
deaths	-0.000991***	0.000488***		-0.000657***
	(0.000276)	(5.19e-05)		(0.000222)
NO2_change_norm		0.176***		
		(0.0287)		
stay_home_duration				-0.00714***
				(0.00226)
povprop				
c.stay_home_duration#c.povprop				
Constant	-24.10***	1.230***	0.0234**	-0.970***
	(0.831)	(0.263)	(0.0112)	(0.165)
Observations	1,429	1,730	1,083	1,085
R-squared	0.591	0.152	0.146	0.347

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
NOTE: aweight = totalpop

The first model uses google mobility data to determine the effect that COVID-19 has had on trends in workplace attendance. Google provided daily data in percent change (from pre-COVID-19 baseline) in workplace attendance for each county. The dependent variable is the average of these daily observations for each county during the COVID-19 outbreak. The negative and statistically significant coefficient on the unemployment trend variable makes sense; counties that saw increases in unemployment, saw decreases in work place attendance. The negative and statistically significant coefficients on deaths and closed business duration is also intuitive; counties with more intense outcomes (more deaths) and longer lockdowns saw lower workplace attendance. These results also demonstrate that counties with higher percentages of Black and Hispanic individuals are positively (and statistically significantly) associated with average workplace attendance. This result seems to reflect the fact that Black and Hispanic workers disproportionately fill roles that cannot be completed from home; 43% of Black and Hispanic workers are employed in service or production. Only about 25% of white workers hold such jobs (according to the American Housing Survey) (New York Times, 2020). This model also shows a positive relationship with higher percentages of a counties population below the poverty line and average workplace attendance. This result is fairly intuitive. Poorer populations still need to attend work more than higher income populations. However, as mentioned above, this is not reflected in poorer populations being positively associated with COVID-19 deaths. One interesting result of this model is that populations with higher death rates from respiratory diseases are positively (and statistically significantly) associated with workplace attendance. Again, this variable was not associated with higher levels of COVID-19 deaths. The positive and statistically significant relationship with intensive care unit hospital beds may indicate that counties that have better health care infrastructure had populations that were more confident in returning to work. However, more data is needed to back up this claim.

The next model examines trends in unemployment. Counties with higher percentages of Black and Hispanic citizens are associated with decreases in unemployment (al-

though the coefficient on Hisp_percent is only significant at the 0.10 level). Again, this seems to be a reflection of the fact that minority populations are disproportionately represented in the service and production industries that have either remained open during the COVID-19 pandemic or are not able to be done remotely. Percentage of uninsured individuals and individuals below the poverty line, duration of business closures, and deaths all had a positive and statistically significant relationship with unemployment. It is also interesting to note that the positive, statistically significant, and largest effect in this model is on the coefficient for the trend in NO_2 concentration. This finding is counterintuitive. We would expect that counties with high levels of unemployment to also exhibit stronger economic shut downs, and be able to adhere to stricter social distancing policies. In theory, this would result in these counties having decreases in pollutant concentrations. However, the data demonstrates the opposite effect. Counties with higher levels of unemployment are positively associated with trends in NO_2 concentrations. This will be examined further in the model focused on trends in NO_2 concentrations.

Before discussing the model on trends in NO_2 concentrations, I would like to highlight the results of the model focusing on the ratio of COVID-19 deaths to confirmed COVIC-19 cases. This model is intended as a way to further model the intensity of the COVID-19 outbreak by county. Simply put, a higher ratio of deaths to cases indicates that a more intense COVID-19 outcome for a county. The results are clear. Counties with higher levels of pre-COVID-19 NO_2 concentrations, older populations, and longer business closures are all associated with more intense COVID-19 outcomes. Clearly, the relationship between business closures and COVID-19 intensity seems to be subject to reverse causality, as areas with more intense outcomes are more likely to keep businesses closed longer. It is interesting to note the coefficient on the variable for percentage of Hispanic population is negative indicating the areas with higher proportions of Hispanic populations are associated with lower death to case ratios and COVID-19 intensity (only significant at the 0.05 percent level). This finding contradicts the finding in the ZINB model and should be cause to further explore this relationship in future research. How-

ever, the negative and statistically significant coefficient on population density provides further evidence that population density is not associated with more intense COVID-19 outcomes.

Finally, for the model on trends in NO_2 concentrations during COVID-19 pandemic, the positive and statistically significant coefficient on the variable for trends in unemployment provides more evidence for the counterintuitive finding that unemployment is positively associated with increases in NO_2 concentrations. However, a measure of unemployment may not be the best way to measure geographically where people are and are not actually physically going to and from their workplace every day. Now more than ever, people who are able to, are working remotely. In order to better asses this, I also included the workplace mobility mean in this regression. The coefficient on this variable indicates a statistically significant and positive relationship between counties where people are still physically going to work and increases in NO_2 concentrations. The coefficient for the percentage of Hispanic population, individuals below the poverty line, and population over 65 were all also positive and statistically significant (hisp_population was only significant to the 0.05 percent level). This demonstrates that counties with higher percentages of people physically going into work, Hispanic populations, individuals below the poverty line, and higher percentages of older people saw increases in their local NO_2 concentrations. It is interesting to note that the coefficient on the percentage of the population that is Black is negative and statistically significant. This indicates that counties with higher percentages of Black individuals actually saw decreases in NO_2 concentrations. The results also show that denser populated counties and counties with higher death rates due to respiratory diseases saw statistically significant decreases in NO_2 concentrations. The same is true for counties with more deaths and longer stay at home orders. This regression indicates that inner city populations (higher population densities), with higher percentages of Black individuals, longer stay at home orders, more intense COVID-19 outcomes, and higher historical death rates of respiratory diseases were the populations expected to see the most significant reductions in NO_2 concentrations. However, it must be said that it is

expected that counties with more intense COVID-19 outcomes and longer stay at home orders would see larger decreases in NO_2 concentrations as we are unable to determine which causes the other (reverse causality).

Initially, this may feel like good news. This seems to indicate that the responses to COVID-19 have themselves caused reductions in pollutant concentrations—which the literature and ZINB models discussed earlier in this paper have shown is a driving force behind COVID-19 intensity—for some of the most vulnerable populations in the country. However, it is important to note that these urban populations that are seeing large reductions in NO_2 concentrations are areas that are associated with some of the highest levels of pollutant concentrations historically. One explanation for this finding may be heterogeneity in the data. In order to help determine whether this was the case, the last piece of analysis I performed was to build a few simple models in order to better determine the effect that the COVID-19 responses have had on the changes in NO_2 concentrations based on socio-demographic characteristics.

This was accomplished by using the same dependent variable for the trends in NO_2 concentrations (standardized) and controlling for the duration of the stay at home order. By interacting the duration of the stay at home order with the variables for the percentage of Black, Hispanic, individuals below the poverty line we should be able to better determine how the responses to COVID-19 have affected the change in NO_2 concentrations for counties with varying levels of these socio-demographic characteristics. The results can be found below.

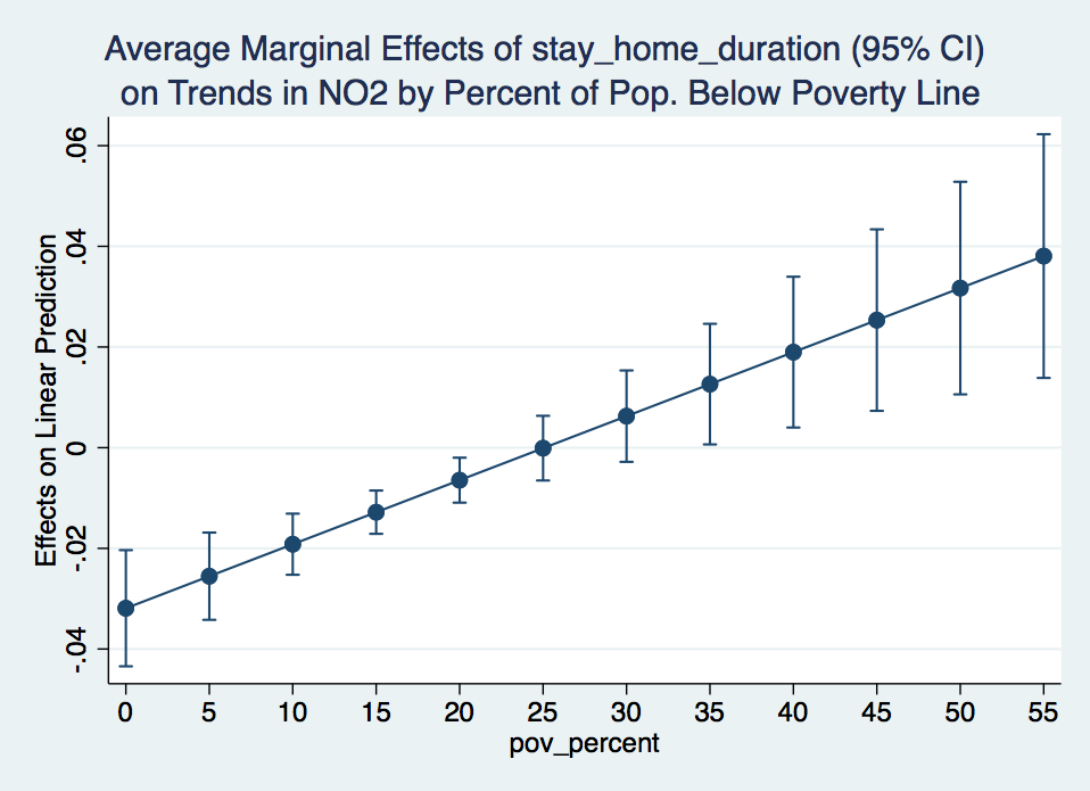
VARIABLES	(1) NO2_trend_std	(2) NO2_trend_std	(3) NO2_trend_std
stay_home_duration	-0.0319***	-0.0176***	-0.0167***
	(0.00588)	(0.00276)	(0.00267)
pov_percent	-0.0415***		
	(0.0119)		
stay_home_duration*pov_percent	0.00127***		
	(0.000321)		
percent_black		-0.0155***	
		(0.00395)	
stay_home_duration*percent_black		0.000303***	
		(0.000112)	
percent_hisp			-0.0297**
			(0.0138)
stay_home_duration*percent_hisp			0.00108***
			(0.000350)
Constant	1.135***	0.758***	0.569***
	(0.231)	(0.109)	(0.103)
Observations	1,212	1,212	1,212
R-squared	0.036	0.043	0.048

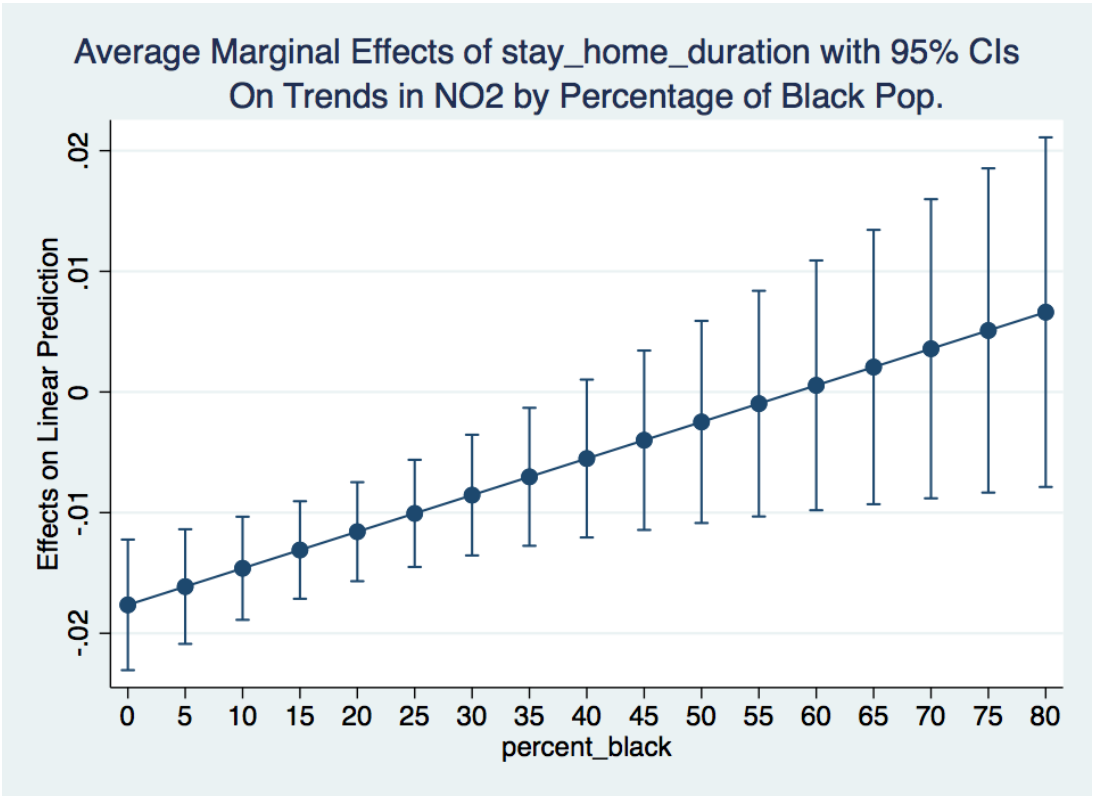
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

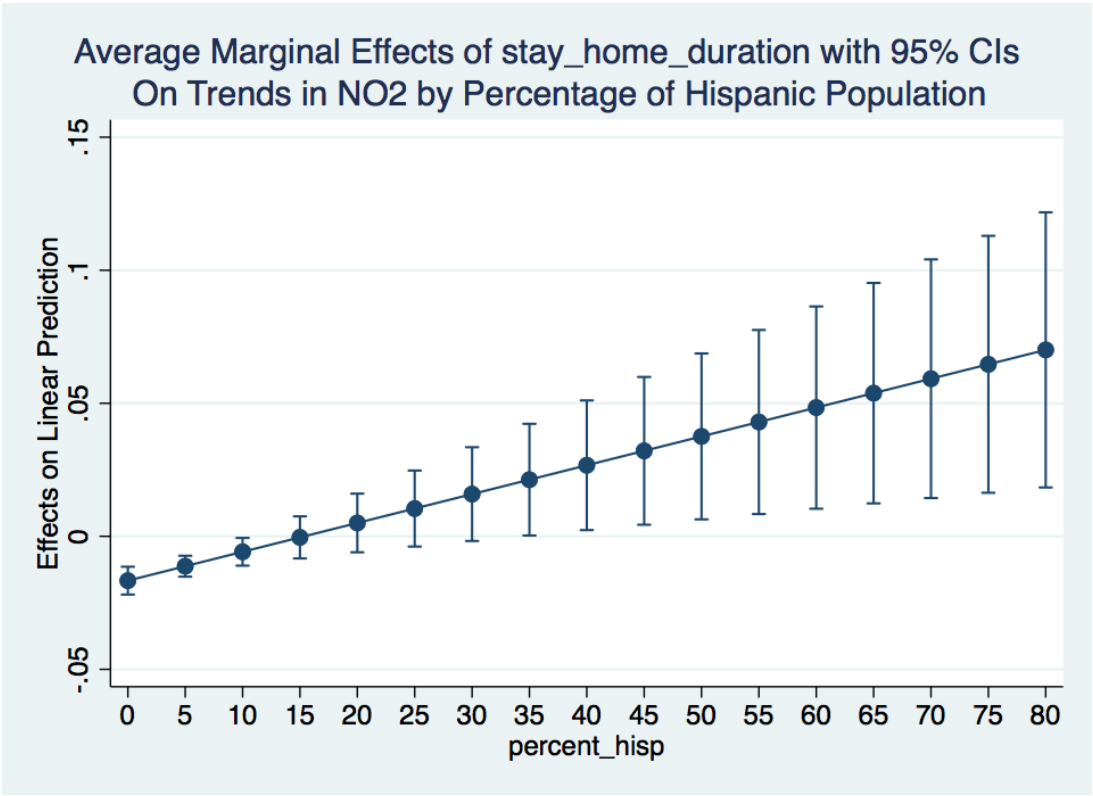
These results indicate that even though the duration of the stay at home order, percentage of population below the poverty line, percentage of Black population, and percentage of Hispanic population all had a negative and statistically significant effect on the trend of NO_2 concentrations on their own, counties with longer stay at home orders and higher percentages of Black, Hispanic, and poor individuals are associated with increases in their NO_2 concentrations.

In order to explore these findings further, I created a plot of the marginal effects of the duration of stay at home orders on the trend in NO_2 concentrations for each of these three interaction models. From the plots below we can see that the negative effects that stay at home orders have had on NO_2 concentrations have been concentrated in counties

with low percentages of people who are below the poverty line, Black, or Hispanic. These results suggest that even though COVID-19 (or more specifically the policy responses to it) has resulted in overall decreases in pollutant concentrations, these improvements in air quality have not been seen in areas that are historically subjected to higher levels of pollutant concentrations. The results for other models presented in this paper suggest that inner city populations are expected to see the most significant reductions in NO_2 concentrations. However, these results demonstrate that poorer, Blacker, and more Hispanic areas are still being disproportionately subjected to higher levels of pollutant concentrations. Based on all of these results, this also appears to be a contributing factor to the disproportionate impacts COVID-19 has had on minority populations in the US.




Average Marginal Effects of stay_home_duration with 95% CIs
On Trends in NO2 by Percentage of Black Pop.
Effects on Linear Prediction
-.02
-.01
0
.01
.02
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
percent_black


Average Marginal Effects of stay_home_duration with 95% CIs
On Trends in NO2 by Percentage of Hispanic Population
Effects on Linear Prediction
-.05
0
.05
.1
.15
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
percent_hisp

5 Conclusion

The data and models presented in this paper indicate that there is a positive and statistically significant relationship between NO_2 concentrations and COVID-19 outcomes. According to the data, a one standard deviation increase in pre-COVID-19 NO_2 concentration is expected to result in 6.5 more COVID-19 deaths. This result is consistent with previous findings that air pollution exposure is linked with worse COVID-19 outcomes. There is also evidence that higher population densities are not necessarily associated with worse COVID-19 outcomes. The results also demonstrated that counties with higher percentages of minority populations in the US are disproportionately affected by COVID-19 independent of the percentage of the county's population that is below the poverty line. A county with 10% more Black people is expected to have 4 more COVID-19 deaths. A county with 10% more Hispanic people is expected to have 3 more COVID-19 deaths. The data indicates that this can be at least partially explained by the fact that minority populations were more likely to continue physically going in to their place of work during the COVID-19 outbreak. And, although responses to COVID-19 (stay at home orders, non-essential business closures, etc.) resulted in decreases in overall NO_2 concentrations, the improvements in air quality were often not present in counties with high percentages of Black, Hispanic, or poor individuals. This indicates that not only have minority populations been disproportionately affected by the immediate impacts of COVID-19 (cases and deaths), but they have also been disproportionately subjected to harmful secondary affects (increases in NO_2 concentrations) leaving them less able to cope with future health risks. Inequality in the US has always been perpetual; COVID-19 appears to be perpet-

uating it even more.

Accurately determining the cause of the impacts discussed in this paper requires access to individual level COVID-19 data where one could accurately control for individual characteristics. This data is not available, and due to the nature of the pandemic, will not be available in the near future. As such, this paper is not, and does not claim to be, estimating a causal relationship. However, the results of this paper indicate that there is an immediate need to continue to enforce regulations on air pollutant concentrations in the US, particularly in areas with historically high levels of air pollution and larger proportions of minority populations. This paper should also serve as justification to continue studying the disproportionate impacts of COVID-19.

References

Abonyi, G. (2020), 'Covid policy lessons for future crises: Social learning framework', *Available at SSRN 3626382* .

Andree, B. P. J. (2020), 'Incidence of covid-19 and connections with air pollution exposure: Evidence from the netherlands'.

Association, A. L. (2001), 'Urban air pollution and health inequities: a workshop report.', *Environmental Health Perspectives* **109**(suppl 3), 357–374.

Association, A. L. et al. (2019), 'American lung association state of the air 2019', *Chicago, IL* .

Barkan, S. E. et al. (2011), *Sociology: Understanding and changing the social world*, Flat World Knowledge, Incorporated.

Cicala, S., Holland, S. P., Mansur, E. T., Muller, N. Z. & Yates, A. J. (2020), Expected health effects of reduced air pollution from covid-19 social distancing, Technical report, National Bureau of Economic Research.

Cole, M. A., Elliott, R. J. & Liu, B. (2020), 'The impact of the wuhan covid-19 lockdown on air pollution and health: A machine learning and augmented synthetic control approach', *Discussion Papers* .

Cole, M., Ozgen, C. & Strobl, E. (2020), 'Air pollution exposure and covid-19'.

Conticini, E., Frediani, B. & Caro, D. (2020), 'Can atmospheric pollution be considered a co-factor in extremely high level of sars-cov-2 lethality in northern italy?', *Environmental pollution* p. 114465.

Coven, J. & Gupta, A. (2020), Disparities in mobility responses to covid-19, Technical report, NYU Stern Working Paper.

Dong, E., Du, H. & Gardner, L. (2020), 'An interactive web-based dashboard to track covid-19 in real time', *The Lancet infectious diseases* **20**(5), 533–534.

Engel-Cox, J. A., Holloman, C. H., Coutant, B. W. & Hoff, R. M. (2004), 'Qualitative and quantitative evaluation of modis satellite sensor data for regional and urban scale air quality', *Atmospheric environment* **38**(16), 2495–2509.

Gates, W. L. (1992), 'An ams continuing series: Global change–amip: The atmospheric model intercomparison project', *Bulletin of the American Meteorological Society* **73**(12), 1962–1970.

Graff Zivin, J. & Neidell, M. (2012), 'The impact of pollution on worker productivity', *American Economic Review* **102**(7), 3652–73.

Hanks, A., Solomon, D. & Weller, C. E. (2018), 'Systematic inequality: How america's structural racism helped create the black-white wealth gap', *Washington: Center for American Progress* .

Hu, Z. (2009), 'Spatial analysis of modis aerosol optical depth, pm 2.5, and chronic coronary heart disease', *International journal of health geographics* **8**(1), 1–10.

Millett, G. A., Jones, A. T., Benkeser, D., Baral, S., Mercer, L., Beyrer, C., Honermann, B., Lankiewicz, E., Mena, L., Crowley, J. S. et al. (2020), 'Assessing differential impacts of covid-19 on black communities', *Annals of Epidemiology* .

Mishel, L. R., Bernstein, J. & Shierholz, H. (2009), *The state of working America: 2008-2009*, Cornell University Press.

Raifman, J., Nocka, K., Jones, D., Bor, J., Lipson, S., Jay, J., Cole, M., Krawczyk, N., Chan, P., Galea, S. et al. (2020), 'Covid-19 us state policy database'.

Rocklöv, J. & Sjödin, H. (2020), 'High population densities catalyse the spread of covid-19', *Journal of travel medicine* **27**(3), taaa038.

Silva, J. M. S., Tenreyro, S. & Windmeijer, F. (2015), 'Testing competing models for non-negative data with many zeros', *Journal of Econometric Methods* **4**(1), 29–46.

Travaglio, M., Yu, Y., Popovic, R., Leal, N. S. & Martins, L. M. (2020), 'Links between air pollution and covid-19 in england', *medRxiv* .

Vuong, Q. H. (1989), 'Likelihood ratio tests for model selection and non-nested hypotheses', *Econometrica: Journal of the Econometric Society* pp. 307–333.

Wilson, P. (2015), 'The misuse of the vuong test for non-nested models to test for zero-inflation', *Economics Letters* **127**, 51–53.

Wu, X., Nethery, R. C., Sabath, B. M., Braun, D. & Dominici, F. (2020), 'Exposure to air pollution and covid-19 mortality in the united states', *medRxiv* .

A Variables

Variable:	Label:	Source:
confirmed	Confirmed COVID-19 cases	JHU CSSE Data
deaths	COVID-19 deaths	JHU CSSE Data
AOD*	2019/2020/difference of standardized Aerosol Optical Depth concentrations	Aerosol Optical Depth 342.5 nm daily 0.25 deg. [OMI OMAEROe v003]
CO*	2019/2020/difference of standardized Carbon Monoxide concentrations	Carbon Monoxide, Mole Fraction in Air (Daytime/Ascending, AIRS-only) daily 1 deg. @1000hPa [AIRS AIRS3STD v006] ppbv
NO2*	2019/2020/difference of standardized Nitrogen Dioxide concentrations	NO2 Tropospheric Column (30% Cloud Screened) (OMNO2d v003)
percent_black	Black population / total population * 100	US Census
percent_hisp	Hispanic population / total population * 100	US Census
povprop	Percent of population living below the federal poverty line (2018)	ACS
beds_icu	Number of intensive care hospital beds	JHU CSSE Data
percent_over_65	Percent of population over age of 65	US Census
popdensity_2018	Population density per square mile (2018)	ACS
rmortality_2018	Crude death rate of respiratory diseases in 2018, # in 100,000	CDC
uninsured_2017	Proportion of population without health insurance in 2017	US Census

unem_trend	Jan-May mean unemployment rate (2020-2019)	BLS
workplace_mobility	Average percent change from baseline of workplace attendance during pandemic	Google mobility data
stay_home_enact	Date state wide stay at home order was enacted	BU CUSP
stay_home_duration	Duration of state wide stay at home order	BU CUSP
closed_NE_business	Date closed non-essential business state wide	BU CUSP
closed_business_duration	Duration of state wide non-essential business closure	BU CUSP

B Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)
VARIABLES	**N**	**mean**	**sd**	**min**	**max**
confirmed	2,757	600.0	2,909	0	80,204
deaths	2,757	31.40	165.9	0	3,780
beds_icu	2,757	32.44	107.8	0	2,352
totalpop	2,757	114,264	344,241	1,605	1.010e+07
_aod2019me	2,742	0.826	0.364	0.0810	3.884
_co2019mea	2,092	144.3	11.92	114.2	187.4
_no2_2019m	2,742	2.054e+15	1.061e+15	1.996e+14	9.790e+15
_so2_2019m	2,742	3.06e-09	1.94e-09	1.48e-10	1.23e-08
_co_2020me	2,232	142.6	9.521	102.0	170.4
_aod_2020m	2,741	0.637	0.320	0.0655	4.607
_no2_2020m	2,742	1.807e+15	8.277e+14	1.878e+14	7.420e+15
_so2_2020m	2,742	3.08e-09	1.98e-09	1.34e-10	1.36e-08
percent change co	2,092	-2.472	11.08	-100	30.84
percent change aod	2,742	-14.51	50.76	-100	906.8
percent change no2	2,742	-8.908	13.54	-54.52	70.92
percent change so2	2,742	0.705	10.24	-46.24	110.3
percent_black	2,757	9.579	14.58	0	87.41
percent_hisp	2,757	9.291	13.72	0	99.07
uninsured_2017	2,757	0.0900	0.0356	0.0280	0.173
rmortality_2018	2,290	112.3	58.61	0	574.4
popdensity_2018	2,757	101.7	614.1	0.0429	27,755
percent_over_65	2,757	14.66	7.686	0	55.60
pov_percent	2,756	15.70	6.216	2.300	55.10
uninsured_per_2017	2,757	9.003	3.559	2.800	17.30
workplace_mean	2,725	-16.10	5.442	-58.33	-3.614
stay_home_duration	1,223	37.50	10.84	24	54
closed_business_duration	1,746	37.97	9.421	19	58
NO2_trend	2,742	-2.465e+14	3.834e+14	-3.530e+15	9.300e+14
NO2_trend_std	2,742	-1.90e-09	1.000	-8.564	3.068
unem_trend	2,757	2.508	1.283	-1.267	9.433